Calhoun: The NPS Institutional Archive

DSpace Repository

Theses and Dissertations | 1. Thesis and Dissertation Collection, all items

2008-06

Integrated data-driven DSS in a laboratory environment

Hargrave, Brian L.

Monterey California. Naval Postgraduate School

http://hdl.handle.net/10945/4133

Downloaded from NPS Archive: Calhoun

http://www.nps.edu/library

Calhoun is the Naval Postgraduate School's public access digital repository for research materials and institutional publications created by the NPS community. Calhoun is named for Professor of Mathematics Guy K. Calhoun, NPS's first appointed -- and published -- scholarly author.

Dudley Knox Library / Naval Postgraduate School
411 Dyer Road / 1 University Circle
Monterey, California USA 93943

NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

INTEGRATED DATA-DRIVEN DSS IN A LABORATORY ENVIRONMENT

by

Brian L. Hargrave

June 2008

Thesis Advisor:	Daniel R. Dolk
Second Reader:	Albert Barreto

Approved for public release; distribution is unlimited

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE		Form Approved OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington DC 20503.

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	June 2008	Master's Thesis

4. TITLE AND SUBTITLE Integrated Data-Driven DSS in a Laboratory Environment	5. FUNDING NUMBERS
6. AUTHOR(S) Brian L Hargrave	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES) N/A	10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

12a. DISTRIBUTION / AVAILABILITY STATEMENT	12b. DISTRIBUTION CODE
Approved for public release; distribution is unlimited	A

13. ABSTRACT (maximum 200 words)

Decision support technologies have remained individualistic as primarily stand alone platforms. The ability to access and integrate a wide range of such technologies in an Integrated Decision Technology Environment (IDTE) can potentially increase a user's ability to create more complex decision support projects. A well-designed IDTE will allow users to identify, learn about, access, execute and integrate disparate decision technologies. Data-Driven DSS provide decision-makers with the capability to store and sort vast amounts of data by leveraging data warehousing and data-mining. These data-oriented decision technologies can assist decision-makers in making better and more informed decisions in shorter durations of time.

This thesis focuses on Data-Driven data mining decision technologies and how they can be integrated into an IDTE. In the process of identifying data mining technology requirements, we first created a simple taxonomy characterized by the four categories of association, classification, clustering, and prediction. We then designed a database schema for storing the requisite data about data mining technologies, and case studies illustrating their use. Finally we designed a simple, yet effective, interface for navigating through the data-driven decision technology universe both at NPS and beyond. SQL commands for populating the various screens of the IDTE interface were provided to show proof of concept.

14. SUBJECT TERMS Data-Driven Technology, Data Mining, IDTE, Integrated Decision Technology Environment, DSS	15. NUMBER OF PAGES 77
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UU

NSN 7540-01-280-5500 Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release; distribution is unlimited

INTEGRATED DATA-DRIVEN DSS IN A LABORATORY ENVIRONMENT

Brian L. Hargrave
Lieutenant, United States Navy
B.S., Old Dominion University, 2001

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN INFORMATION TECHNOLOGY MANAGEMENT

from the

NAVAL POSTGRADUATE SCHOOL
June 2008

Author: Brian Hargrave

Approved by: Daniel Dolk
Thesis Advisor

Albert “Buddy” Barreto
Second Reader

Dan Boger
Chairman, Department of Information Sciences

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

Decision support technologies have remained individualistic as primarily stand alone platforms. The ability to access and integrate a wide range of such technologies in an Integrated Decision Technology Environment (IDTE) can potentially increase a user's ability to create more complex decision support projects. A well-designed IDTE will allow users to identify, learn about, access, execute and integrate disparate decision technologies.

Data-Driven DSS provide decision-makers with the capability to store and sort vast amounts of data by leveraging data warehousing and data-mining. These data-oriented decision technologies can assist decision-makers in making better and more informed decisions in shorter durations of time.

This thesis focuses on Data-Driven data mining decision technologies and how they can be integrated into an IDTE. In the process of identifying data mining technology requirements, we first created a simple taxonomy characterized by the four categories of association, classification, clustering, and prediction. We then designed a database schema for storing the requisite data about data mining technologies, and case studies illustrating their use. Finally we designed a simple, yet effective, interface for navigating through the data-driven decision technology universe both at NPS and beyond. SQL commands for populating the various screens of the IDTE interface were provided to show proof of concept.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

LIST OF FIGURES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

ACKNOWLEDGMENTS

I would like to thank my wife Mary and my daughters Katharine, Kristina, and Kimberly for their love and support because they made this experience worthwhile. I would also like to thank Professor Dan Dolk for his direction and guidance. Without him, this process could not have been possible. Finally to my friends, your friendship here at the Naval Postgraduate School will follow me for the rest of my life, fair winds and following seas.

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

A. PROBLEM CHARACTERIZATION

A decision support system is a system under the control of one or more decision makers that assists in the activity of decision making by providing an organized set of tools intended to impose structure on portions of the decision-making situation and to improve the ultimate effectiveness of the decision outcome.[1]

DSS (DSS) have become increasingly important in today's society as computing technology moves from strictly operational functions to problem-solving activities. The amount of data that needs to be processed and transformed in order to make informed decisions has increased in concert with contemporary computers' ability to store and process ever more vast collections of data. The ability to "make sense" out of exponentially increasing amounts of data makes the use of DSS, and associated decision technologies, a more vital cog in the information landscape.

Data-driven DSS provide decision-makers with the capability to store and sort through vast amounts of data by leveraging data warehousing and data-mining. These data-oriented decision technologies can assist decision-makers in making better and more informed decisions in shorter durations of time. This thesis will focus on data-driven technologies, particularly data mining, and how they can be integrated with other decision technologies.

[1] George M. Marakas, Decision Support Systems In The 21st Century, 2d ed. (Upper Saddle River: Prentice Hall, 2003), 4.

B. MOTIVATION

DSS are interactive computer-based systems or subsystems intended to help decision makers use communications technologies, data, documents, knowledge and/or models to identify and solve problems, complete decision process tasks, and make decisions. DSS assist in the activity of decision-making by providing an organized set of tools to impose structure on portions of the decision-making process. Currently, integrated DSS environments with a portfolio of available decision technologies are largely nonexistent. While DSS tools are frequently used by both staff and students at NPS, there is no single portal or access point to available software. An integrated decision technology environment (IDTE) would allocate these available tools in a central environment for use. Further, capabilities would be provided to integrate various technologies in support of specific applications. Additionally, the IDTE will host self-tutorials, publications, and other web-content that will benefit users. This research will focus on data-driven decision technologies that will be included into the integrated Decision Support Laboratory environment.

C. METHODOLOGY

The approach will be to conduct a literature survey of data-driven DSS. The next step will be to identify and categorize different data-driven technologies. The methodology will continue with the development of a data-driven decision technology taxonomy that will be used in designing an interface for the IDTE. This taxonomy will be

enhanced via construction of an equivalent database schema for storing relevant information about data mining technologies and associated case studies. This database schema will drive the design of a conceptual user interface for the IDTE. An overarching use case will be generated describing the usage of the IDTL in the data-driven decision technology mode.

Current resources from the areas of Data-driven technologies, data-mining, data warehousing, and web-based DSS are integrated in order to develop a conceptual framework for the IDTE.

D. SCOPE AND LIMITATIONS

This thesis will identify an approach to the preliminary design and development of an IDTE with special emphasis on data mining technologies as a subset of data-driven DSS. A working model will be designed but not implemented. This thesis will provide a conceptual view of the database models to be incorporated into future research and development of the IDTE.

E. OUTLINE OF THESIS

Chapter II of this thesis describes data mining, data warehousing, and knowledge discovery technologies, including an ontology in the form of a simple taxonomy. Chapter III presents a conceptual data mining technology schema to be utilized in the development of the IDTE. Chapter IV presents a conceptual user interface design for the IDTE based upon the database schema and following a use case scenario.

Chapter V provides a summary and recommendations for future enhancements of the IDTE.

II. DATA WAREHOUSES AND DATA MINING

A. GENERAL

In this chapter, we examine two key concepts critical to the research. We will conduct a brief overview of data warehouses and data mining as they pertain to DSS. From this review, we will develop a taxonomy of data mining followed by a survey of available software at Naval Postgraduate School (NPS). The purpose of developing this taxonomy is to provide a framework for eventual users of the IDTE to access and use data mining software systems effectively.

B. DATA WAREHOUSES

A data warehouse is a specific type of database designed to support managerial decision-making. A database is simply "a collection of interrelated data organized to meet the needs and structure of an organization and can be used by more than one person for more than one application.[2] A database by itself may not provide the user with useful decision-making knowledge, particularly if it is operational in nature. A standalone, operational database primarily provides a place for data to be stored in a structured way until needed for report generation and/or queries. It allows

[2] Efraim Turban and Jay E. Aronson, Decision Support Systems And Intelligent Systems, 5th ed. (Upper Saddle River: Prentice Hall, 1998), 80.

the user to access, view, and change data as required, mostly with the objective of maintaining timely and accurate data.

A data warehouse, on the other hand, differs in its objectives. Data Warehouse definitions vary widely, but a serviceable version is "data warehousing combines data from multiple databases or data sources into a large database for the purpose of providing more extensive information retrieval and data analysis."[3] A data warehouse is the backbone of a DSS. As shown in Figure 1, all information is taken from or put into the data warehouse for the primary purpose of knowledge discovery.

[3] Shon Harris, All In One CISSP Exam Guide, 3rd ed. (New York: McGraw-Hill, 2005), 830.

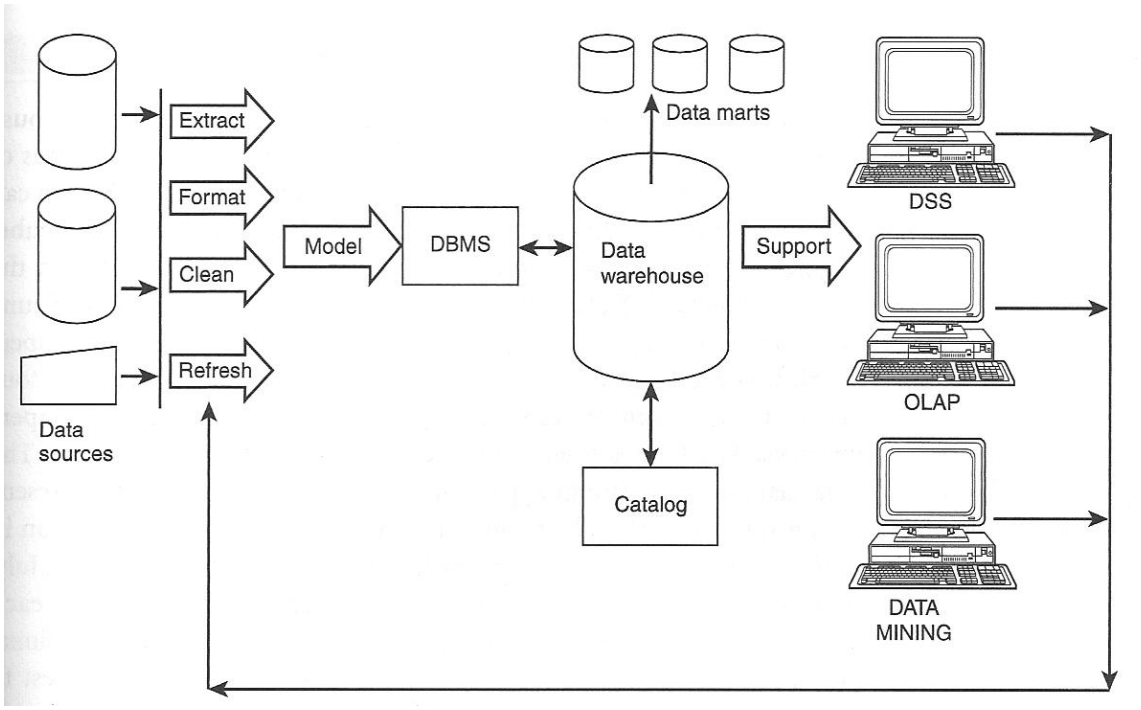

Figure 1. Data Warehouse Architecture.[4]

The Architecture also illustrates how the data is taken from multiple sources, formatted, cleaned (checked for integrity, validity, duplicity, and relevance), and refreshed as it is incorporated into the warehouse. Data Marts, as shown in Figure 1, are subsets of the data warehouse. The data mart can be a small snapshot of the larger data warehouse or it can be a repository of specialized information about a particular area of data. An example of a data mart is one of an enterprise that has a data warehouse but creates data marts that separate the different departments or regions of the company. This

[4] Catherine M. Ricardo, Databases Illuminated (Sudbury: Jones and Bartlett, 2004), 735.

provides the benefits of working on a smaller scale while limiting the damage a mistake will permeate through the warehouse.

A data warehouse serves several functions beyond that of simple storage of data. The data warehouse is most useful when utilizing the data retrieval and analytical functions. As shown in Figure 1 previously discussed, these functions include DSS (DSS) using querying and reporting tools, On-Line Analytical Processing (OLAP), and data mining. The main differentiators of a data warehouse is that, via OLAP, it provides the ability to look at data multi-dimensionally, and via data mining to discover patterns across historical data.

Querying and reporting are the simplest tasks for the data warehouse. The query is what it sounds like. It is a question that the user generates to the data warehouse. The questions that are asked will be specific in nature. An example of a simple query that a user might ask a database is to list the Navy Personnel who are male. This will work as long as the database has a table of Navy Personnel and one of the attributes listed in that table is the gender of the service member. This will return a list of all Navy personnel that are male. The queries can be more complex. These more complex queries look across several different tables in the database. An example of a complex query might be the same query as before but with additional restrictions, such as a list of Navy personnel who are male and participating in the Thrift Savings Plan (TSP). The query would ask the database to search the table Navy Personnel for male personnel and search the table TSP for

the male Navy personnel. There would have to be a link that connects the information from one table to the other. Generally, this is some identifiable unique element that will be included in the tables. For the above example, a good element would be the Navy Personnel's social security number (SSN). Figure 2 shows a simplified view of how the tables in the database example would be connected.

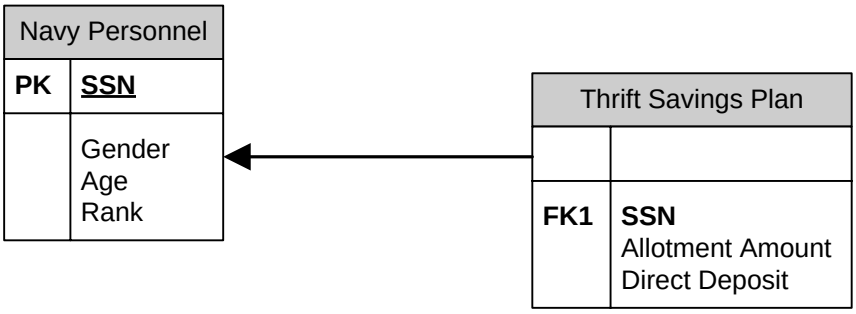

Figure 2. Sample Entity Relationship Diagram.

This pattern of setting up the databases and the tables in the databases continues for even more complex queries that search across multitudinous tables with numerous attributes each.

Online Analytical Processing (OLAP) provides the user with the capability to conduct multidimensional analysis of data to include data modeling, trend analysis, and other complex calculations. OLAP allows the user to conduct ad hoc queries in multiple dimensions providing flexibility and speed to the search in databases that contain a very large and complex amount of data. OLAP software is best utilized when incorporated into the database as it is created. The data will be formatted and structured to support rapid, multi-dimensional queries. The star schema is a popular

framework implemented in a relational database. Figure 3 shows how a simple star schema would store data in the database.

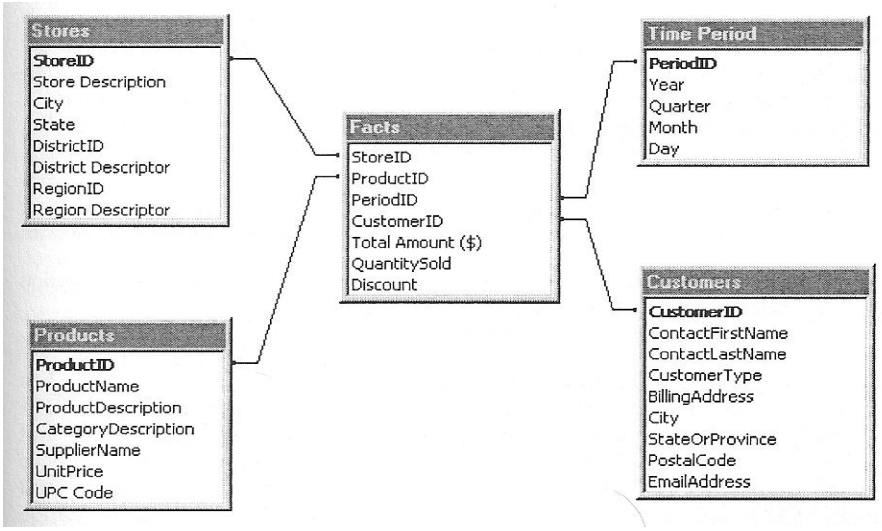

Figure 3. Star Schema Diagram.[5]

As shown in Figure 3, a star schema will have a central table called a fact table that contains attributes common to the other tables. The dimension tables are then linked to the central table by foreign keys.[6] The attributes on the fact table quantifies the data represented by the dimension tables. This allows the combination of data from multiple tables that can be aggregated while maintaining their meaning. The dimension tables in contrast describe the data organized in the fact table. It presents unique keys connecting the tables together. This provides the ability to

[5] Daniel J. Power, Decision Support Systems: Concepts And Resources For Managers (Westport: Quorum Books, 2002), 127.

[6] Ibid.

quickly perform analysis on very large datasets.[7] OLAP analysis becomes increasingly beneficial when combined with data mining for analysis.

C. DATA MINING

Data Mining (also called knowledge data discovery, KDD) is best defined as the use of automated data analysis techniques to find previously unknown connections between data stored in a data warehouse. The knowledge gained from utilizing these techniques and finding the hidden relationships and patterns associated with the data can help the user gain a competitive advantage over those who do not.[8] Data mining tools are also used to automate the process of predicting outcomes from information in large databases.

The data mining techniques can vary from company to company as they adjust the techniques to fit their particular needs. The basic techniques that are focused on in this research are the common categories that are currently in use: 1) classification, 2) association, 3) prediction, and 4) clustering.[9]

1. Classification

Classification is a data mining approach that develops rules that determine if the item belongs to a particular set of data. These rules are generally simplistic in nature.

---

[7] George M. Marakas, Decision Support Systems In The 21st Century, 2d ed. (Upper Saddle River: Prentice Hall, 2003), 332-333.

[8] Daniel J. Power, Decision Support Systems: Concepts And Resources For Managers (Westport: Quorum Books, 2002), 143.

They often follow a simple logic sequence such as IF the data meet X parameters, THEN it will be placed with other data that meet the same parameters. These rules can be modified to provide a specific percentage of accuracy. These modified conditions traditionally are a) exact rule, b) strong rule, and c) probabilistic rule.[10]

a. Exact Rule

The exact rule gives one hundred percent probability that the data will meet the parameters of the class that it is placed into. It explicitly fulfills the preconditions of the IF-THEN statement. There are no exceptions to the rule that are allowed.[11]

b. Strong Rule

The strong rule takes a range of exceptions into account when admitting data into a certain classification. This rule allows for a range between ninety and one hundred percent probability. The data element in this case does not have to exactly match.[12]

c. Probabilistic Rule

The probabilistic rule creates a subclass that classifies the data based on a measured probability. It does

---

[9] George M. Marakas, Decision Support Systems In The 21st Century, 2d ed. (Upper Saddle River: Prentice Hall, 2003), 333.

[10] George M. Marakas, Decision Support Systems In The 21st Century, 2d ed. (Upper Saddle River: Prentice Hall, 2003), 334.

[11] Ibid.

[12] Ibid.

this by relating the conditional probability P (THEN|IF) to the probability P (THEN).[13]

The data mining technique utilizing classification requires the analyst to set up the classes beforehand in which to place the data. The specific questions that the company need answered are determined early and the classes are set up accordingly to answer them.

An example utilizing classification would be a video rental stores customer rental history. The classes would be set up based on several attributes of a typical customer base, which would likely include categories such as age, gender, movie type, and membership type. This would allow the video store to determine the movies most likely to be rented and by whom. Based on the data in the classes, they will then be able to predict given a pre-determined probability how many copies of a new movie to purchase and who they will expect to rent based on the type of movie.

2. Association

Association rules relates one set of data to others to find patterns of events that happen concurrently. This relationship is denoted by {event A} → {event B}. This shows that when event A takes place event B is also likely to take place. While the above mentioned relationship indicates each event as a singular entity, event A and event B could each represent a group of events that are associated with each other. Similar to the rules that further define

[13] George M. Marakas, Decision Support Systems In The 21st Century, 2d ed. (Upper Saddle River: Prentice Hall, 2003), 334.

classification, association has two measures represented by percentages that lend a numerical value to the interaction between events: support and confidence.[14]

a. Support

Support as it pertains to data mining association rules is the percentage of a data set that contain all the elements that make up event A and all the elements that make up event B.[15] In other words, that percentage of the data records for which Event(A) -> Event(B).

b. Confidence

Confidence is closely related to support except it takes event A as the data set and finds how many times event B present.[16]

For example, an association rule may be generated that when a video renter rents a children's movie, he is likely to buy candy. This would be expressed as {children's movie} → {candy}. Knowing this association would benefit the video store because it would allow them, say, to place the candy near the children's movie section in the video store.

Using this example, the support measure would be the ratio of those who both Rent(Movie) & Purchase(Candy) to all

[14] Catherine M. Ricardo, Databases Illuminated (Sudbury: Jones and Bartlett, 2004), 748.

[15] Catherine M. Ricardo, Databases Illuminated (Sudbury: Jones and Bartlett, 2004), 748.

[16] Ibid.

those who Rent(Movie). If, out of 1,000 movie rentals, 100 renters also bought candy, then the support would be 10%.

Now assume that out of the 1,000 movie rentals, 250 of them were children's movie rentals and that from the 250children's movie rentals, 50 of them sold candy at the same transaction. The Confidence level would then be 20 percent to the association rule.

The association technique to data mining can be useful in finding less obvious connections between seemingly diverse objects. An example of such a connection that proves the value of the technique is the association that men who buy diapers frequently purchase beer as well.[17] Such a connection is less intuitive but can reap many benefits for company advertising and product placement.

3. Prediction

Prediction follows the association analysis pattern to determine relationships between independent events. Prediction differs in the fact that it utilizes the completion of an event and past history to predict the next event that is most likely to occur. It forms a predictive chain of events that can be expected to take place based on past data that has been collected and analyzed. Prediction follows a timeline of events that only advances when the previous event completes. As data collects on specific event chains, a percentage can be calculated to indicate the probabilities that succeeding events will follow the

[17] George M. Marakas, Decision Support Systems In The 21st Century, 2d ed. (Upper Saddle River: Prentice Hall, 2003), 334.

sequence.[18] The popular models used for prediction are decision trees and neural networks.

a. Decision Trees

Decision trees are sequences of events that are derived from the data by using logic statements and rules to partition the data sets into increasingly smaller categories. This drilling down of the data sets continues until a user defined stopping point is reached, or the partitioning occurs at the simplest level. Much like classification, the decision nodes are IF-THEN statements that split the data sets into multiple branches. The completed tree will form a complete path of events based on the data.[19] This will allow a predictive relationship to form, connecting the input to a probabilistic end state.

b. Neural Networks

Neural networks are data models that are created to adapt to new data as it is integrated into the model. Neural networks find patterns from past data to predict outcomes.[20] A typical neural network follows a three step process. First it predicts an outcome, and then compares the prediction to the actual outcome. The final step is an adjustment of the model, depending on whether there was a match or not. This is the simulated learning process.[21]

---

[18] Catherine M. Ricardo, Databases Illuminated (Sudbury: Jones and Bartlett, 2004), 749.

[19] George M. Marakas, Decision Support Systems In The 21st Century, 2d ed. (Upper Saddle River: Prentice Hall, 2003), 336.

[20] Daniel J. Power, Decision Support Systems: Concepts And Resources For Managers (Westport: Quorum Books, 2002), 151.

[21] Ibid., 152.

Following the three step process, the neural network represents a human reaction and decision making process. A human will adapt the decisions that they make as new information and data becomes available. The neural network is developed to mimic that human response to new data.

A prediction example would follow the purchasing habits of a customer captured at a hardware store. It can be predicted that when a customer buys drywall, he will most likely purchase nails the next visit. After numerous customers and visits, the data from their purchases will be collected and the sequence of events can be predicted with a certain probability of actually occurring.

4. Clustering

Clustering techniques are related to the classification technique in that they group data based on similar characteristics. These groupings are called clusters. Where classification determines the groups before analysis, clustering determines the grouping during analysis based on attributes and metrics determined by the user. This technique is especially useful when the groupings cannot be predefined.[22]

Clustering techniques typically implement certain algorithms to define their cluster parameters. These algorithms are; a) Hierarchical and b) Partitioning.

[22] George M. Marakas, Decision Support Systems In The 21st Century, 2d ed. (Upper Saddle River: Prentice Hall, 2003), 335.

a. Hierarchical

Hierarchical clustering techniques allow the user to analyze nested data at different levels into which data fields are sub-divided. It does this by creating a diagram similar to a decision tree that is made up of clusters of data. Hierarchical clustering can either be agglomerative or divisive.

Agglomerative hierarchical clustering starts with each point of data as an individual cluster. The pair of clusters closest together then combines into one cluster. This pattern continues for an undetermined amount of iterations until a pre-determined number of clusters are left or until only one cluster remains.

Divisive hierarchical clustering takes an opposite approach to agglomerative clustering. It starts with a single cluster of all the data points. The cluster then splits into two separate clusters. The cluster will continue to split until a pre-determined number of clusters are reached.[23] Typically, only a single cluster will split at successive iterations.

b. Partitioning

Clustering using the partitioning technique places data into cluster sets. These cluster sets do not overlap and each data point will only be a member of one cluster.[24]

---

[23] Pang-Ning Tan, Michael Steinbach, and Vipin Kumar, Introduction To Data Mining (Massachusetts: Addison-Wesley, 2006), 515.

[24] Ibid., 492.

There are two popular methods to determine how the clusters are organized, K-medoids method and K-means method.

The k-medoids method takes a selected number of data points and uses them to represent the centers of clusters. Data points are then added to the clusters based on the proximity to the initial data points.[25]

The k-means method is the more popular partitioning cluster method. The clusters are grouped around the centroids. The centroid will be the geometric center of the cluster of data points. The initial centroids are chosen by the user. The surrounding points are then added to the clusters. The centroid location is then updated based on the geometric mean of the points. The data points are then adjusted to the clusters to which they are more closely located. The centroid location then is calculated based on the new cluster of data points. The iterations continue until the centroid locations no longer change.[26]

Clustering is often utilized in analyzing medical data. For example, a clustering of medical data relating symptoms to diagnosed diseases can be useful to find hidden or seemingly unrelated symptoms that were previously undetected.

D. CONCEPT TAXONOMY IN DATA MINING

A data mining taxonomy can be developed based on the descriptions and definitions defined above. The taxonomy

---

[25] Pavel Berkhin, Survey Of Clustering Data Mining Techniques (San Jose: Accrue Software, 2002), 15.

[26] Pang-Ning Tan, Michael Steinbach, and Vipin Kumar, Introduction To Data Mining (Massachusetts: Addison-Wesley, 2006), 497.

developed places the different data mining techniques into an ordered representation of the relationships.

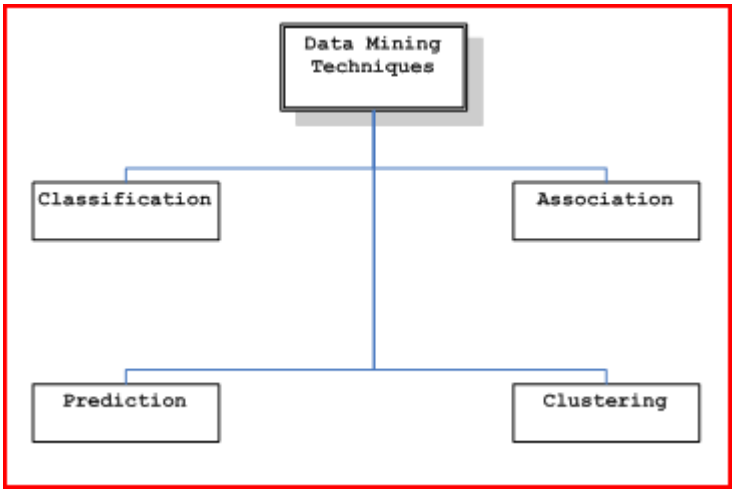

Figure 4. Taxonomy of Data Mining.

The taxonomy shown in Figure 4 will be utilized and followed in the next chapter. This taxonomy gives a base for further research to be implemented into an IDTE combining the data-driven technologies described with other DSS technologies, such as model-driven and knowledge-driven.

E. SURVEY OF AVAILABLE SOFTWARE

A survey of available data mining software at the Naval Postgraduate School found two main software programs in use: 1) SPSS Clementine and 2) Megaputer PolyAnalyst.

1. SPSS Clementine

SPSS Clementine provides a graphical user interface that shows the steps in the data mining process as they are being used. Clementine offers the user the ability to utilize the techniques of data mining mentioned in this chapter. It is designed to allow the user to choose between

Clustering, Classification, Association, and Prediction.[27] Figure 5 shows the Clementine interface and start-up screen.

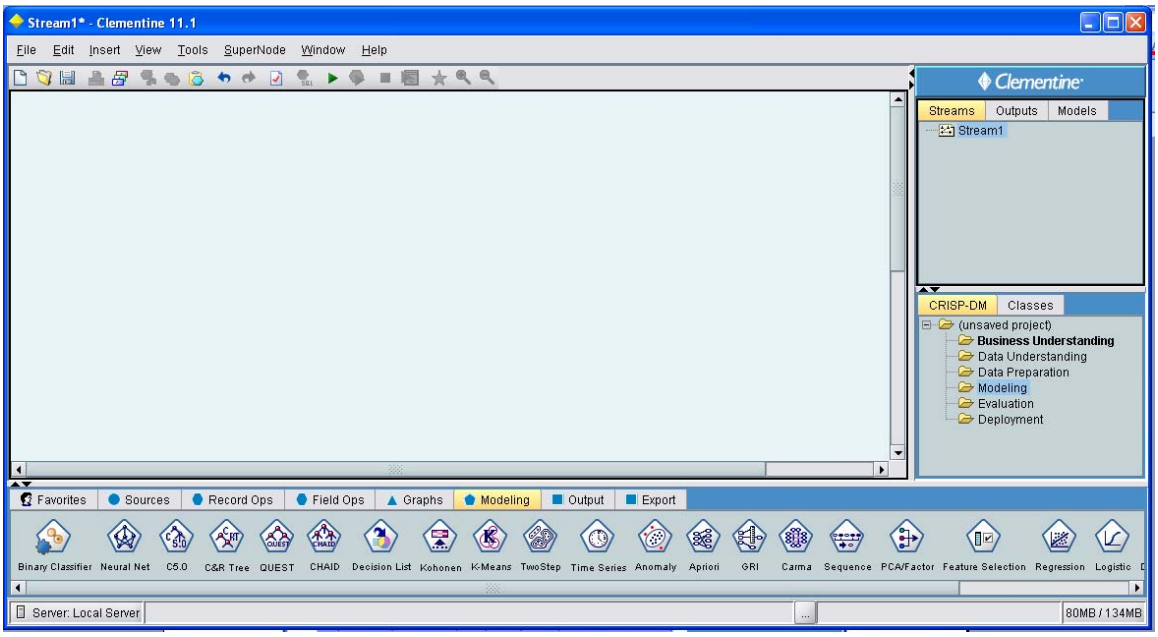

Figure 5. Clementine Start-up Screen.

2. Megaputer PolyAnalyst

Megaputer PolyAnalyst is a robust software package that utilizes several data mining techniques for knowledge discovery. The techniques that PolyAnalyst can perform are: Classification, Clustering, Association, and Prediction which includes Pattern Learning and Trend Analysis.[28]

---

[27] SPSS Clementine 11.1 Specification Brochure. 2005. Online. Internet. 12 Feb. 2008. Available from http://www.spss.com/pdfs/CLM11SPClr.pdf.

[28] Metaputer PolyAnalyst Brochure. 2007. Online. Internet. 12 Feb 2008. Available from http://www.megaputer.com/polyanalyst.php.

Figure 6 shows a snapshot of the features available in the software package.

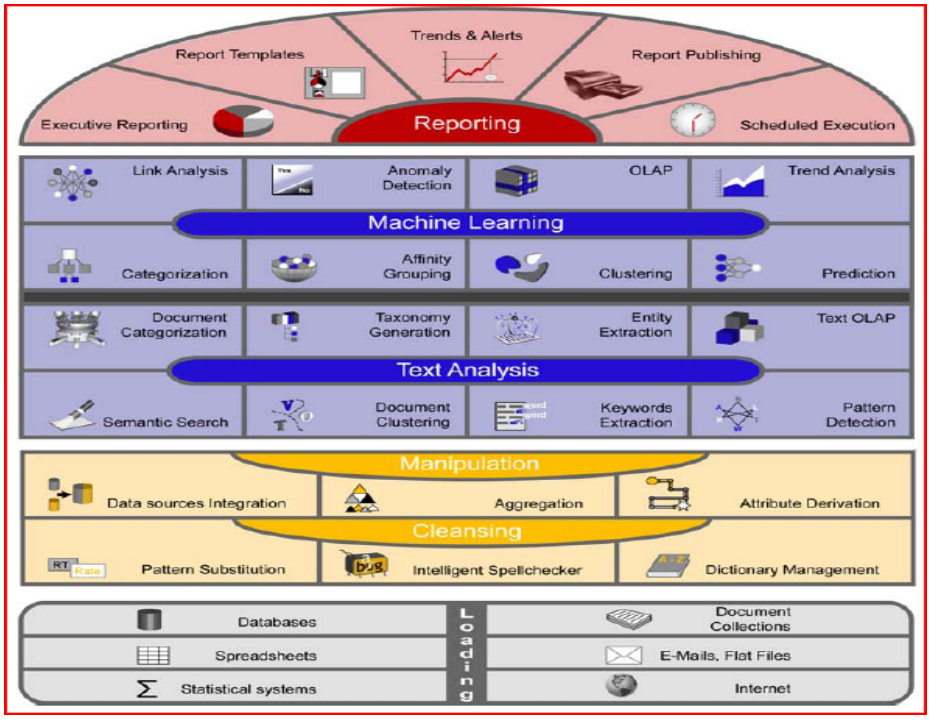

Figure 6. PolyAnalyst Features Diagram.[29]

Megaputer PolyAnalyst combines these techniques into an easy to read graphical interface and provides custom reports on the results. Figure 7 shows the user interface and startup screen for PolyAnalyst. From this startup screen the user can navigate to an existing project, create a new project, or go to a tutorial section that has fifteen different tutorials. These tutorials will walk the user through a variety of techniques from a simple introduction to data mining and PolyAnalyst operation to more in-depth tutorials.

[29] Metaputer PolyAnalyst Brochure. 2007. Online. Internet. 12 Feb 2008. Available from http://www.megaputer.com/polyanalyst.php.

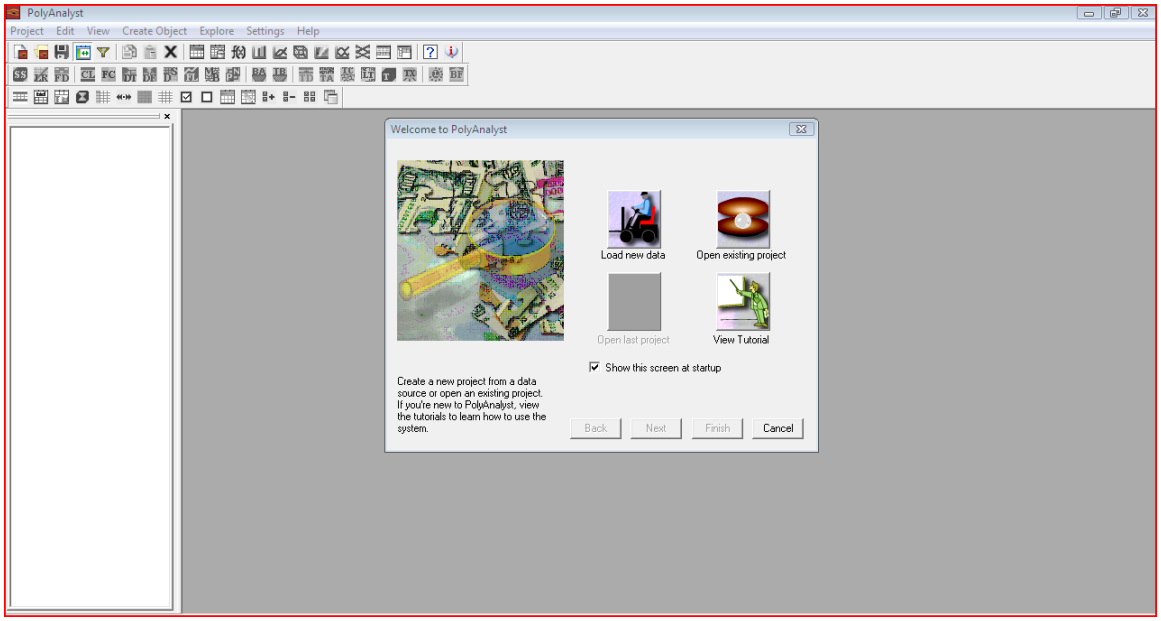

Figure 7. PolyAnalyst Start-up Screen.

F. SUMMARY

In this chapter we reviewed the essential data warehousing and data mining structures and technologies. A formalized taxonomy of data mining technologies was presented and we concluded with a survey of the available software at NPS, SPSS Clementine and Megaputer PolyAnalyst.

In the next chapter (Chapter III – Data Mining Schema) we will provide a concept Decision Technology and Data Mining Schema to be utilized in the development of the IDTE.

THIS PAGE INTENTIONALLY LEFT BLANK

III. DATA MINING SCHEMA

A. INTRODUCTION

In this chapter, we provide a concept Decision Technology and Data Mining Schema to be utilized in the development of the IDTE. The tables needed for the operation of the IDTE will be described and the relationships between the tables will be identified. These schemas are not all inclusive and leave the ability to add additional tables and relationships as necessary to increase the range of decision technologies as they become available to the user. This chapter and thesis focus on data-driven decision technologies and as such will not develop in detail the other decision technologies in the schemas or in the relationships beyond the first level. These tables and their relationships will be the basis for the creation of the concept user interface described in the next chapter.

B. DATA MINING SCHEMA

The data mining schema was created utilizing the Tabledesigner program. The following is a brief description on how to read the Tabledesigner diagram and interpret the symbols contained within.

Objects: An object can be regarded as the equivalent of an entity, and typically will be represented by at least one

corresponding database table in a database. Figure 8 shows the Decision Technology object.

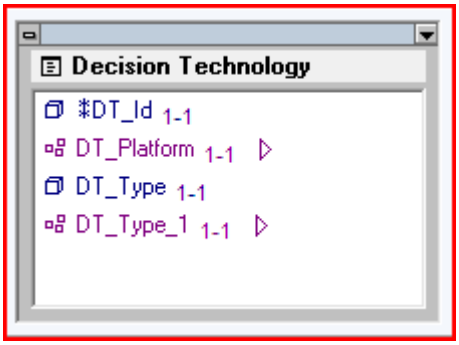

Figure 8. Decision Technology Object.

Each object in Tabledesigner may contain one or more of the following attributes: simple attribute, group attribute and object attribute. All attributes are assigned cardinalities which can be zero-to-one (0.1), one-to-one (1.1), zero-to-many (0.N), or one-to-many (1.N).

Simple attributes: A simple attribute is equivalent to a data item representing a single piece of information. In terms of database design, data items represent columns in a database table. Referring to Figure 8, 

would be one of the data items for the object Decision_Technology.

Group attributes: A group attribute is a container within an object that collects items that are conceptually related. It does not appear anywhere in a data table, rather it visually makes it easier for the user to see and relate

appropriate data items. Figure 9 shows the group attribute DT_Platform which groups several simple attributes pertaining to the platform.

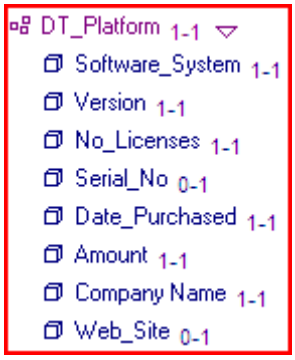

Figure 9. Data Group Example.

Object attributes: Object attributes form the mechanism for specifying relationships between two objects. These relationships are represented by including an object attribute in another object which requires the reverse process as well. Figure 10 shows this relationship, as Data Mining_DT has an attribute in the DM_Technique Object and vice versa. This represents a relationship between the two.

The details of that relationship are indicated by the subscript numbers or cardinality that follows the attribute.

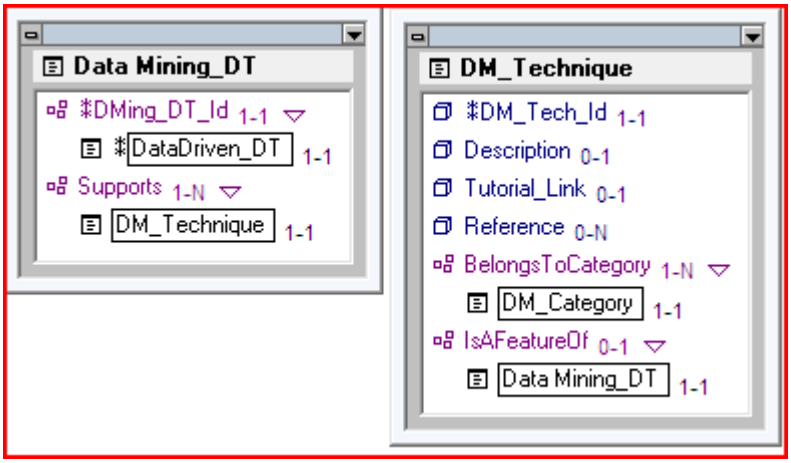

Figure 10. Object Link Example.

Cardinalities allow the creator to determine the number of allowable instances in each direction of a two-way relationship that is required.

One-to-one relationship: This would be indicated in the schema by the Maximum Allowed (the second subscript number) property set to one on both sides of the relationship. This means that one instance can be associated with only one other instance of a related item.

One-to-many relationship: This is a relationship where one instance can be associated with many other instances, and many instances can be associated to a single instance. This is shown by one instance having its Maximum Allowed property set to one and linked to other instances that have their Maximum Allowed property set to N for indicating numerous.

Many-to-many relationship: This is a relationship where many different instances can be associated with many other

different instances. This is shown by having the Maximum Allowed property set to N on both sides of a two-way relationship.

Figure 11 is an initial data mining schema emphasizing data-driven decision technology process.

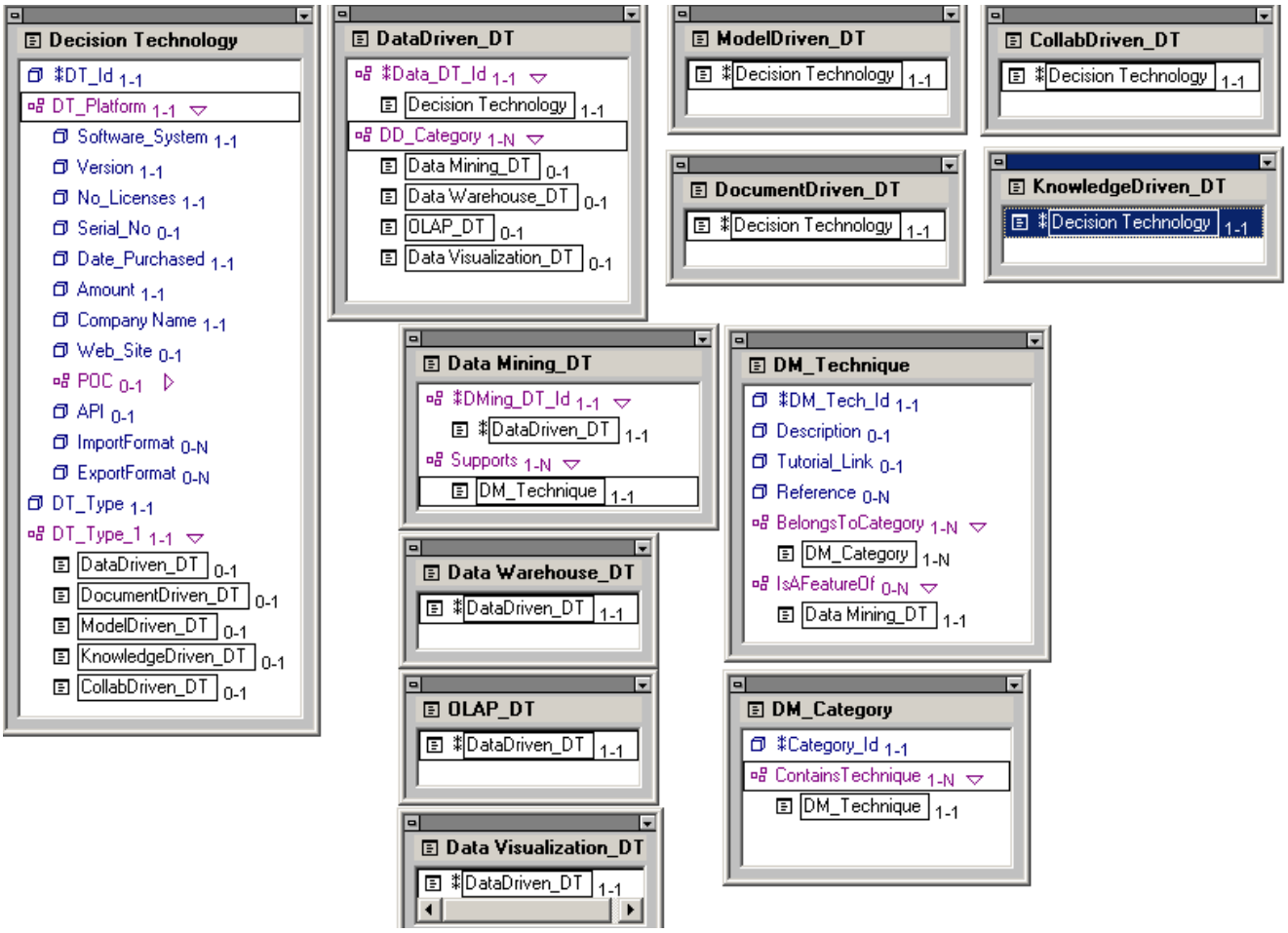

Figure 11. Data Mining Schema.

This schema can be represented as a taxonomy in Windows Explorer format which would look similar to Figure 12.

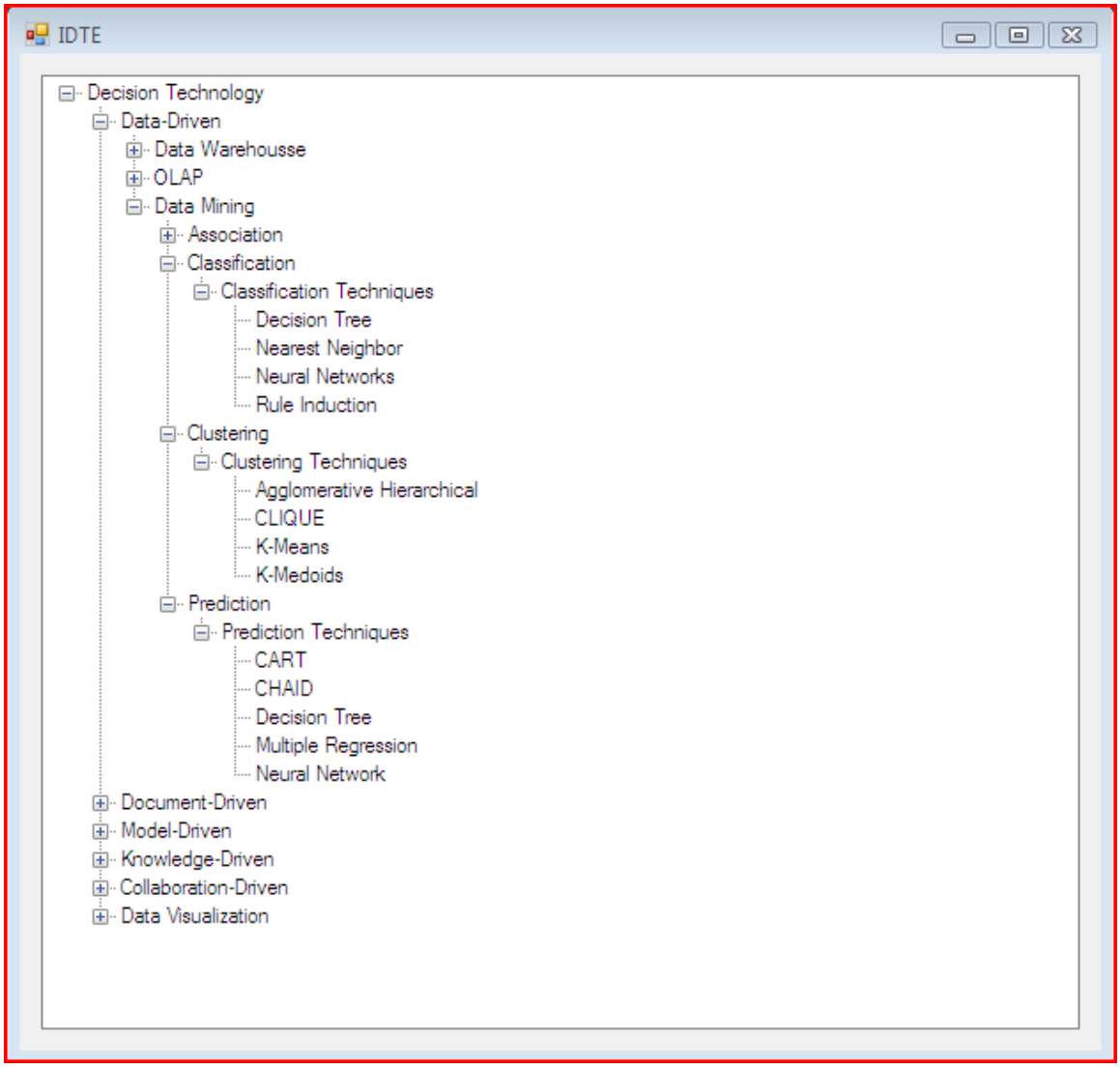

Figure 12. Windows Explorer Taxonomy.

Table 1 outlines the attributes and provides a brief description, data type and example of the items included in the data mining schema.

Attribute	**Description**	**Attribute Type**	**Example**
DECISION TECHNOLOGY	A single software system in the IDTE	Object	Megaputer, Clementine
DT_Platform	Salient information about the software, some but not all attributes of which are hinted at here	Group	Megaputer
:	Attributes listed		
DT_Type	Decision technology type	Group	Data-driven, Model-driven, etc
DataDriven_DT	One instance for each data-driven software DT; inherits the key from the corresponding Decision Technology instance		
DD_Category	Group attribute shows the generalization hierarchy another level deep with data-driven DTs	Group	Data Warehouse, OLAP, Data Mining, and/or Data Visualization
DataMining_DT	One instance for each Data Mining data-driven software (e.g., Megaputer); inherits the key from the corresponding DataDriven_DT instance		
DM_Technique	One instance for		Decision_Tree

	each data mining technique		
DM_Category	One instance for each of four data mining categories		Classification, Clustering, Association and Prediction
Description → Reference	These attributes refer to the documentation available for this software decision technology; these are not intended to be inclusive; additional attributes may be desirable		

Table 1. Attribute Table.

Table 2 outlines relationships and provides a brief description, and example for the data mining schema.

Relationship	**Description**	**Example**
Supports 1-N.DM_Technique 1-1	shows one-to-many relationship that an instance of DataDriven_DT may have many corresponding instances of DM_Technique	
IsAFeature 0-N .DataMining_DT 1-1	shows zero-to-many relationship between DM_Technique and DataMining_DT	the *DecisionTree* technique may appear as a feature in both *Megaputer* and *Clementine*
BelongsToCategory	one-to-many	the *DecisionTree*

1-N . DM_Category 1-1	relationship showing that an instance of a DM_Technique may belong to at least one and perhaps many instances of DM_Category	technique may appear as a feature in both *Megaputer* and *Clementine*
ContainsTechnique 1-N.DM_Technique 1-1	shows one-to-many relationship between DM_Category and DM_Technique	the category *Prediction* may include *Regression* and *NeuralNetworks* techniques as well as many others

Table 2. Relationship Table.

C. DATA MINING SCHEMA EXTENDED TO INCLUDE CASE STUDY

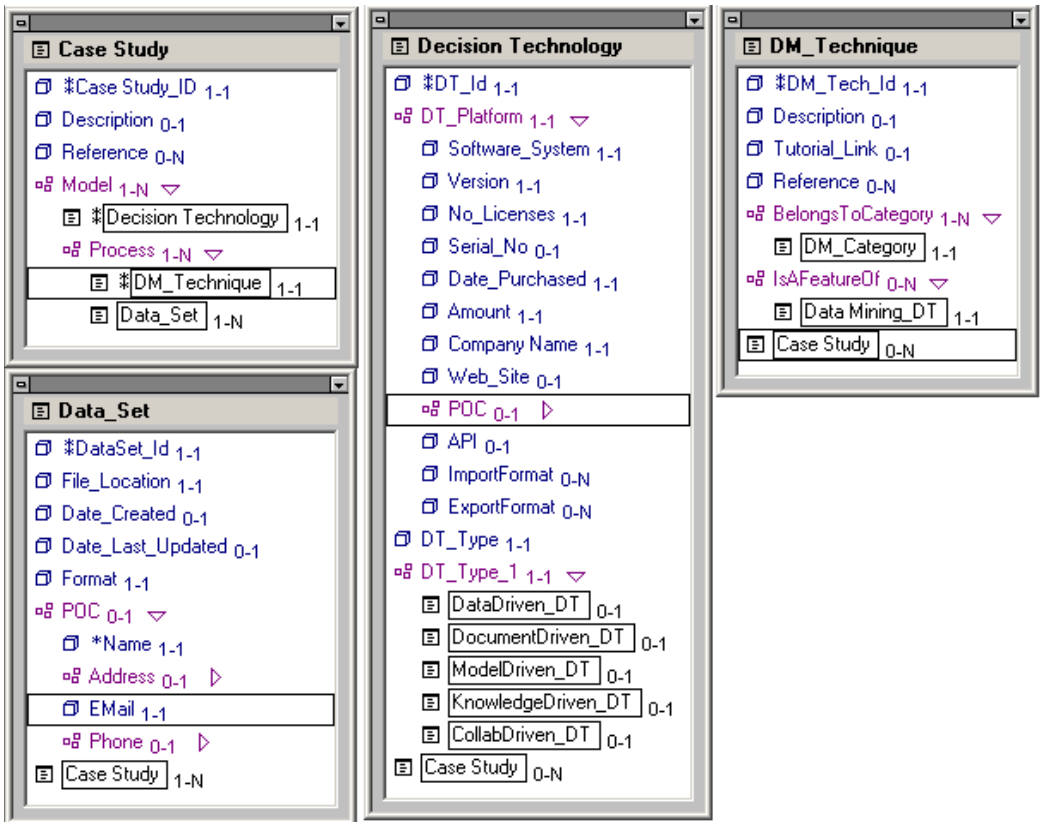

Figure 13. Schema including Case Study.

The schema in Figure 13 has been enhanced to include applications in the form of the Case Study object. New or modified objects are described below in Table 3.

Attribute	Description	Attribute Type	Relation ship	Description
Case Study	This object is intended to capture relevant information about any data mining applications which have been developed and saved in			

	the IDTE			
Model 1-N	Intended to capture the fact that an application may have included one or more Decision Technologies each of which may use one or more DM_Techniques and Data_Sets	Group		
DM_Technique:			*Case Study* *0 – N*	A DM_Technique may have appeared in 0, 1, or more instances of Case Study
Data_Set	This object is meant to capture instances of reusable datasets in the IDTE. Note that this object will need considerably more attributes to fully describe each data set		*Case Study* *0 – N*	A Data_Set may be used in 0, 1, or more instances of Case Study

Table 3. Case Study Attribute/Relationship Table.

D. DATABASE TABLES

The Tabledesigner program allows the user to create a database from the model/schema created in the program. Based on the schema depicted in Figure 8, a database was created in Microsoft Access. From that database, relationship

diagrams and database tables were produced allowing the ability to created user queries.

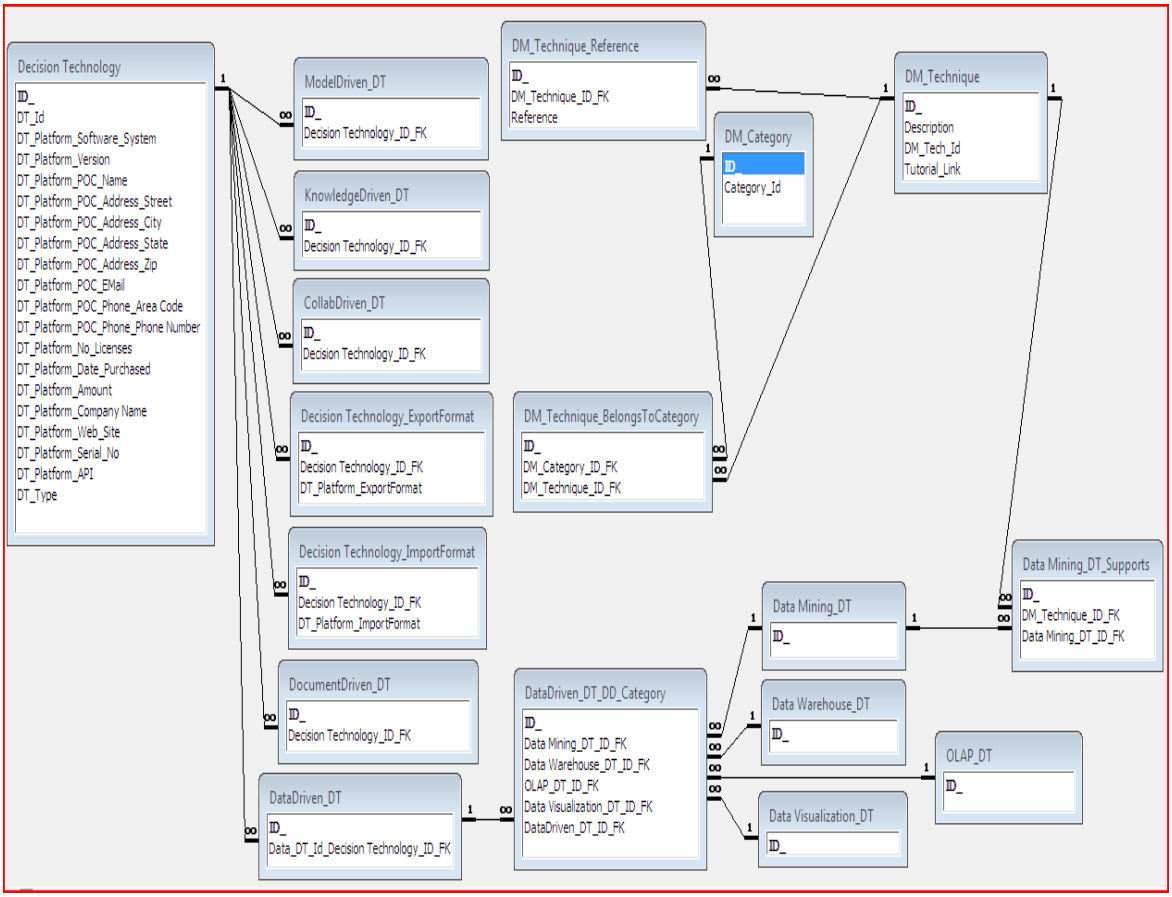

Figure 14. Microsoft Access Entity Relationship Diagram.

From the diagram presented in Figure 14, we can develop the database tables. These database tables, examples shown

in Figures 15, 16, and 17, are the basis for the SQL queries described in section E. The tables are shown with data records.

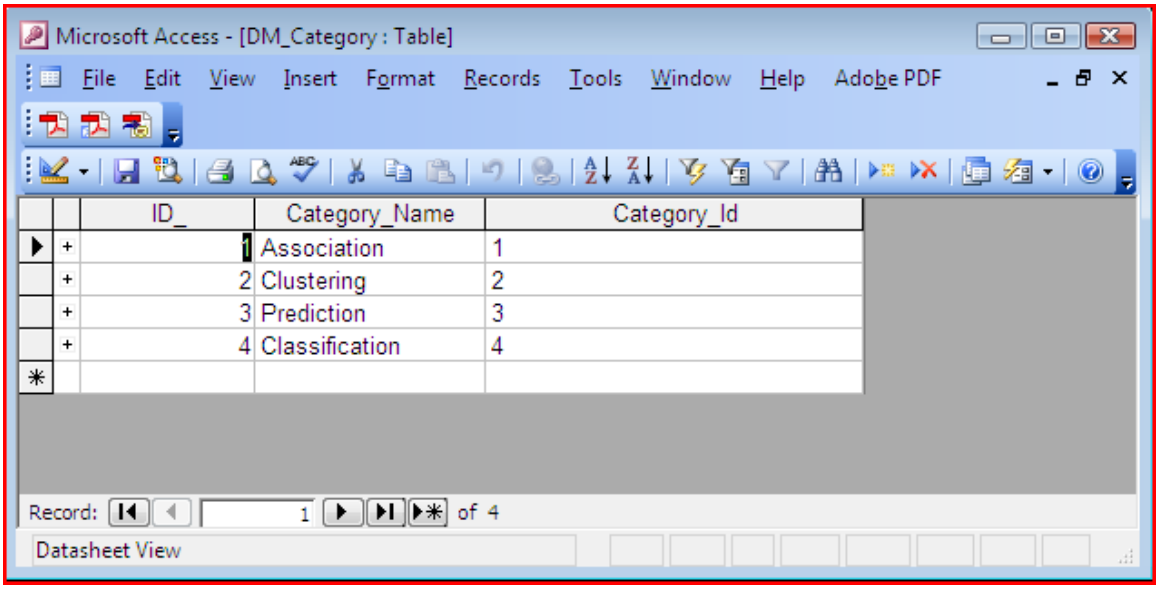

Figure 15. DM_Category Table.

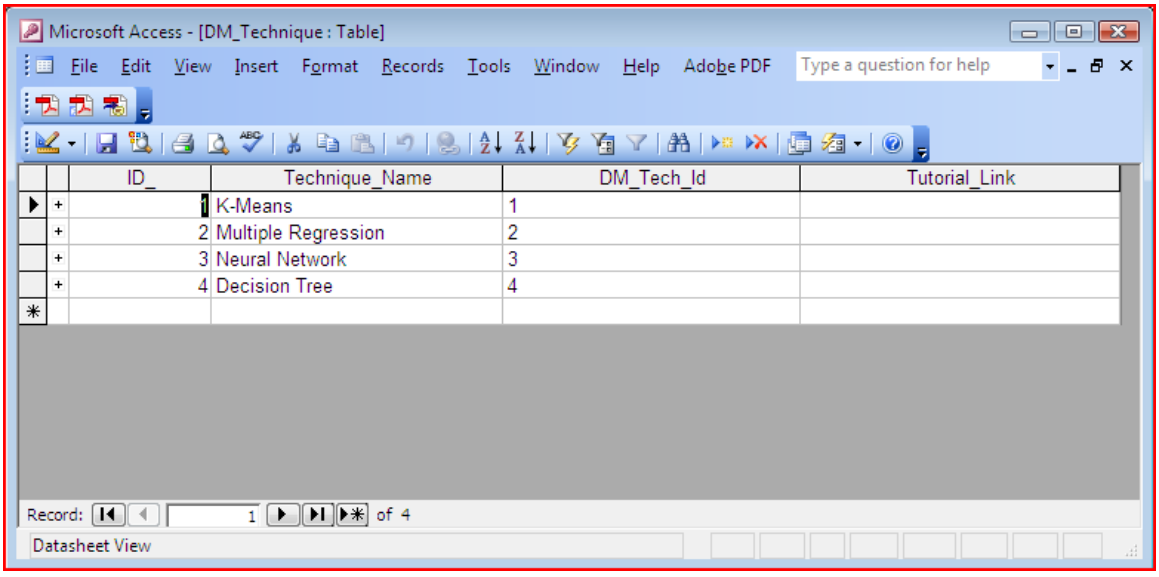

Figure 16. DM_Technique Table.

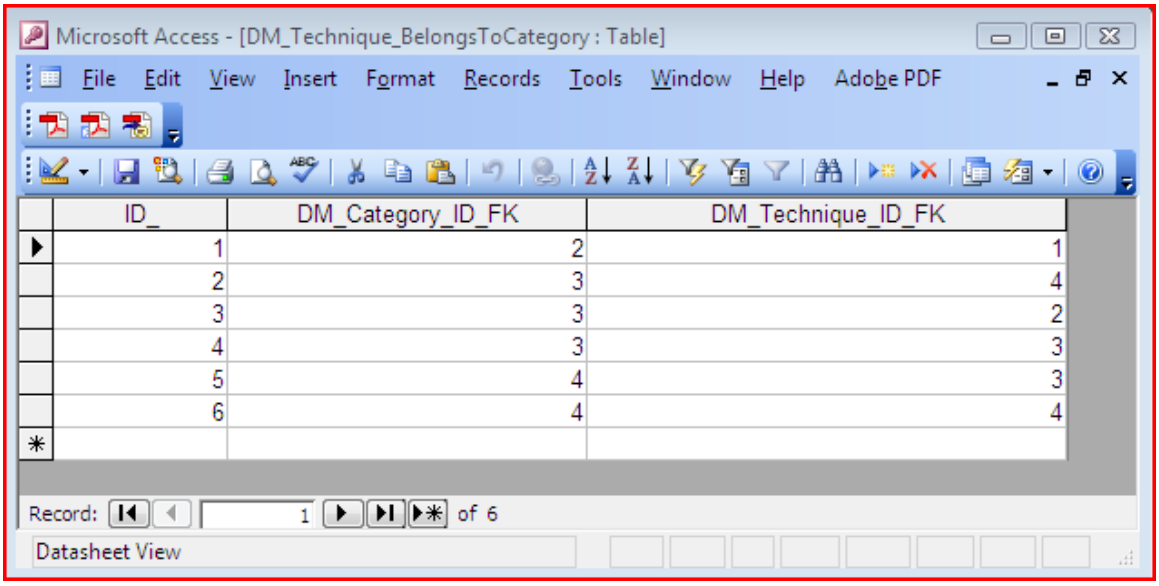

Figure 17. DM_Technique_BelongsToCategory Table.

Fiqure 15 shows the records of the different data mining categories. Figure 16 shows a selection of data mining techniques. Figure 17 shows the many-to-many relationship between the categories and the techniques.

E. QUERIES

The SQL query language allows the user to retrieve data from the database tables by specifying what data they are looking for. The SQL query language specifies the SELECT statement as the means of creating the queries the user wants answered. The general format for an SQL query follows:

```
SELECT column_name(s)
FROM table_name
```

The SELECT determines what columns that the user wants to search for the data. The FROM determines what database table the user wants to look for the data in. The results are

displayed in a result table. An example of a general SQL query from the database table provided in Figure 16 would be:

```
SELECT DM_Technique.Technique_Name
FROM DM_Technique;
```

This SQL query would return a result table like the one in Table 4. The columns would be filled with the data contained in the database, for example a listing of the data mining techniques.

	Technique_Name
▶	K-Means
	Multiple Regression
	Neural Network
	Decision Tree
*	

Table 4. Data Mining Technique Result Table.

Additionally, the queries can be expanded to join two or more tables into a single query to combine and organize the data to the user's specifications. An example of a more advances SQL query that joins the data mining categories to the data mining techniques to produce a result table showing what techniques are part of which categories would look like the following:

```
SELECT DM_Category.Category_Name,
DM_Technique.Technique_Name
FROM DM_Category INNER JOIN (DM_Technique INNER JOIN
DM_Technique_BelongsToCategory ON DM_Technique.ID_ =
DM_Technique_BelongsToCategory.DM_Technique_ID_FK) ON
DM_Category.ID_ =
DM_Technique_BelongsToCategory.DM_Category_ID_FK;
```

This SQL query would return a result table like the one in Table 5.

Category_Name	Technique_Name
Clustering	K-Means
Prediction	Decision Tree
Prediction	Multiple Regression
Prediction	Neural Network
Classification	Neural Network
Classification	Decision Tree

Table 5. Data Mining Category/Technique Result Table.

There are numerous SQL Queries that can be created that can answer most questions the user may have regarding what techniques are associated with which category, which platform can perform which techniques, or what data driven technology is appropriate for a given case study.

F. SUMMARY

In this chapter we reviewed the initial decision technology and data mining schemas, the database tables, and how SQL queries created from those database tables would look. These include the adaptation of case studies into the schema. The schemas allow for further development into other decision technologies beyond the data driven/data mining technology.

In the next chapter (Chapter IV – User Interface) we will utilize the schema and database tables described as the foundation for building a conceptual user interface for the IDTE.

IV. USER INTERFACE

A. INTRODUCTION

In this chapter, we design a conceptual user interface for the IDTE. A hypothetical use case will motivate a typical IDTE user through the screens of the IDTE that will guide the user to one of two existing programs, either SPSS Clementine or Megaputer PolyAnalyst, for performing a decision tree analysis on a dataset. Screen captures will be shown of the conceptual IDTE User Interface that supports this process.

A user in general may want to perform several different tasks or look-ups from the Data-Driven / Data Mining segment of the IDTE. The Use Case illustrates someone looking for a specific technique, decision tree analysis, and navigating to a software environment which supports that technique. A user may also want to utilize the IDTE to scan existing Use Case Studies and Applications for guidance in how to set up a specific problem. The user can access the IDTE and its associated links to tutorials to familiarize themselves with existing software platforms. The IDTE can thus be used as a resource or teaching aid in the Data-Driven / Data Mining field. Users can also browse to identify what techniques and/or software platforms exist in the overall data mining universe.

B. IDTE FLOWCHART

The use case for this chapter is a generic description of the process needed to analyze a data set using a decision

tree. Once a platform has been identified and selected, the IDTE will transfer the user to the desired program environment.

The first step in the development of the user interface is to create a navigation flowchart that shows how the user will progress through the different screens of the IDTE. The flowchart in Figure 18 follows only the path that a user would follow to select a software package for applying a decision tree technique for analysis. The other paths are not shown to their completion but would have similar options available based on associated attributes.

Figure 18. IDTE Flowchart.

The flowchart does not show all possible interconnections in that each page can in principle be connected to every other page. The user will be able to link

to any page based on a selection from a Windows Explorer type tree available on each page viewed. This is explained further in the next section.

C. IDTE USER INTERFACE

The conceptual IDTE was created with Microsoft Visual Studio 2005 using Visual Basic. The choice of Visual Studio was based on familiarity with Microsoft programs and user interfaces. Consistency and compatibility with the installed Windows platforms prevalent at Naval Postgraduate School computer laboratories was an additional factor.

Each screen provides simple buttons that are labeled to indicate what will happen when they are selected. Each screen will have the same look and feel to the user, including the familiar minimize, maximize, and close buttons in the right top hand corner.

Figure 19 is the welcome screen that allows the user to become accustomed to the format that the IDTE will follow. The screens will be designed with four distinct areas of functionality. The first is the title area, centered at the top of the screen, which identifies the current screen. Below and to the right of the title area is a window with a basic definition of the screen. To the left of the definition area is the Windows Explorer navigation area. This area allows the user to expand nodes in a familiar tree fashion displaying the different screens available for selection. By clicking on the “+” sign the tree will be expanded to the next lower level and by clicking on the “–“ sign, the tree will collapse the level up to the next higher category. Users always have the option to click on any title

in this section to proceed directly to the desired page after they have become more familiar with the type of decision technology they require, as shown below. The final area (the button area) is located in the lower right hand side of each screen. Each screen other than the welcome screen will have the five buttons shown in Figure 2 and a “Back” button that will return to the previous screen. The buttons from top to bottom are: Platforms, References, Case Studies, Tutorials/Demos, and Google Search.

The “Platforms” button functionality depends upon which page is current. When selected, it will navigate the user to a screen that lists the platforms available to meet the requirements for that particular screen.

The “References” button opens a separate window with a list of references that the user can access to gain further insight into Decision Technologies. This selection is also context sensitive, showing only relevant references for the focus of the current screen.

The “Case Studies” button opens a separate window with a list of applicable case studies based on the current screen. Case studies can be added to a database and categorized based on technology, category, technique, and platform.

The “Tutorial/Demos” button opens the tutorial supplied by the platform software relevant to the current screen. These will also be opened up in separate windows so the user can toggle between the IDTE and the tutorial.

The final button is a “Google Search” button. This button opens a separate window to allow the user to Google search as required for ad hoc information.

Figure 19. IDTE Welcome Screen.

Once users enter the environment they will be greeted by a Decision Technologies screen (Figure 20). The Decision Technologies are listed on the left of the screen. If the user does not know which technology is appropriate for his purpose, he can utilize a “mouse over” feature to hover the mouse over the words in the Windows Explorer tree area to bring up a pop-up window with the definition of that Technology from the Title section of that page. This will allow the user to make a decision about what selection to

navigate to without a trial and error approach on which technology to follow. We assume for this use case that we want to select the Data Driven Technology.

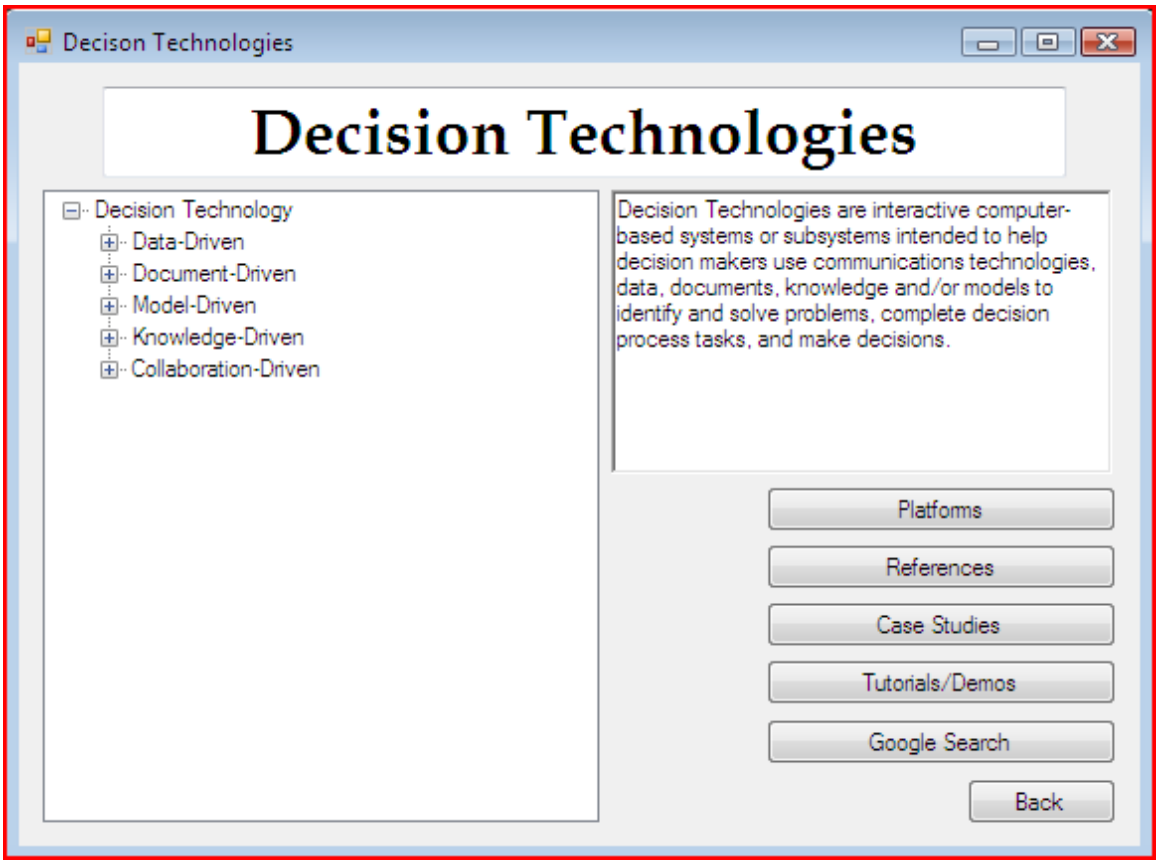

Figure 20. Decision Technology Screen.

Figure 21 shows the Data-Driven Technology Screen, which contains the four subject areas of Data Warehousing, OLAP, Data Mining and Data Visualization. At this point, we can see how the interface aligns with the Windows Explorer paradigm and so will proceed to the Data Mining area which is the focus of this work.

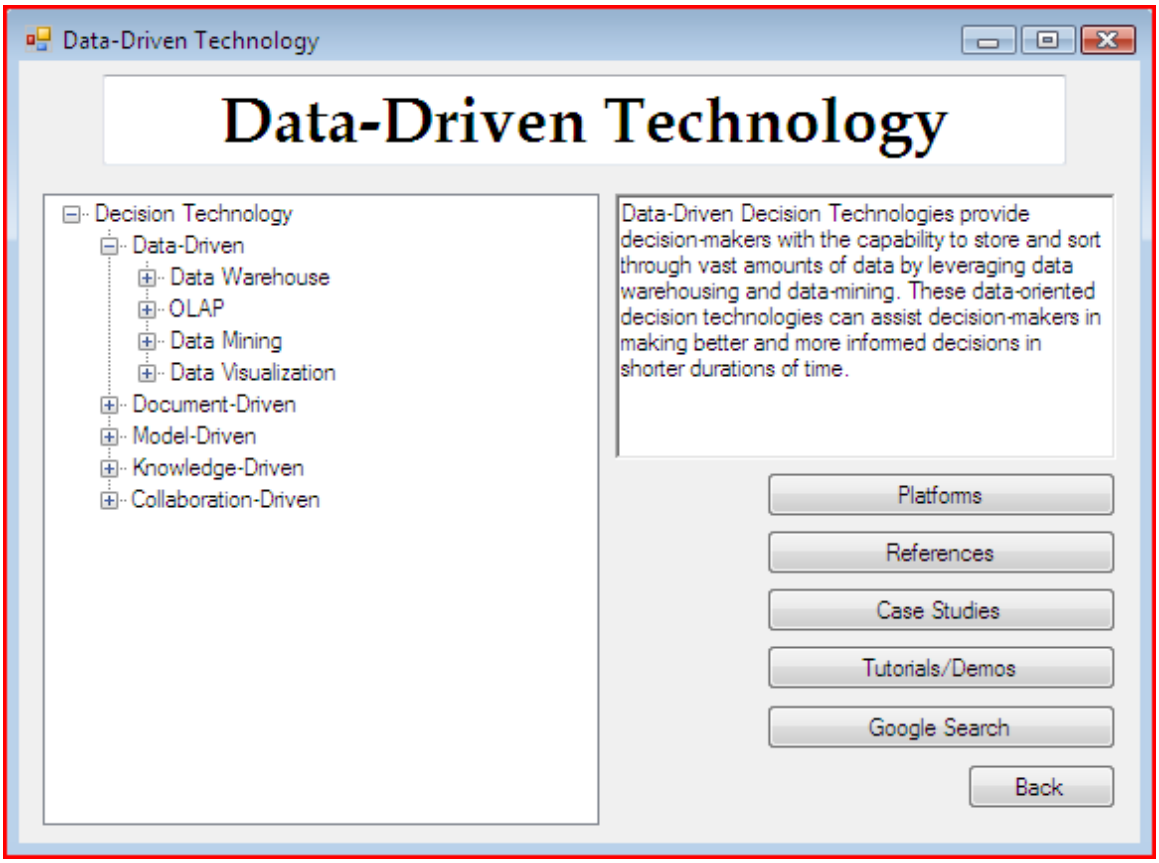

Figure 21. Data-Driven Technology Screen.

The use case we are considering is one where an analyst is searching for a platform which can create decision trees from datasets. Exploding the Data Mining option yields the screen in Figure 22.

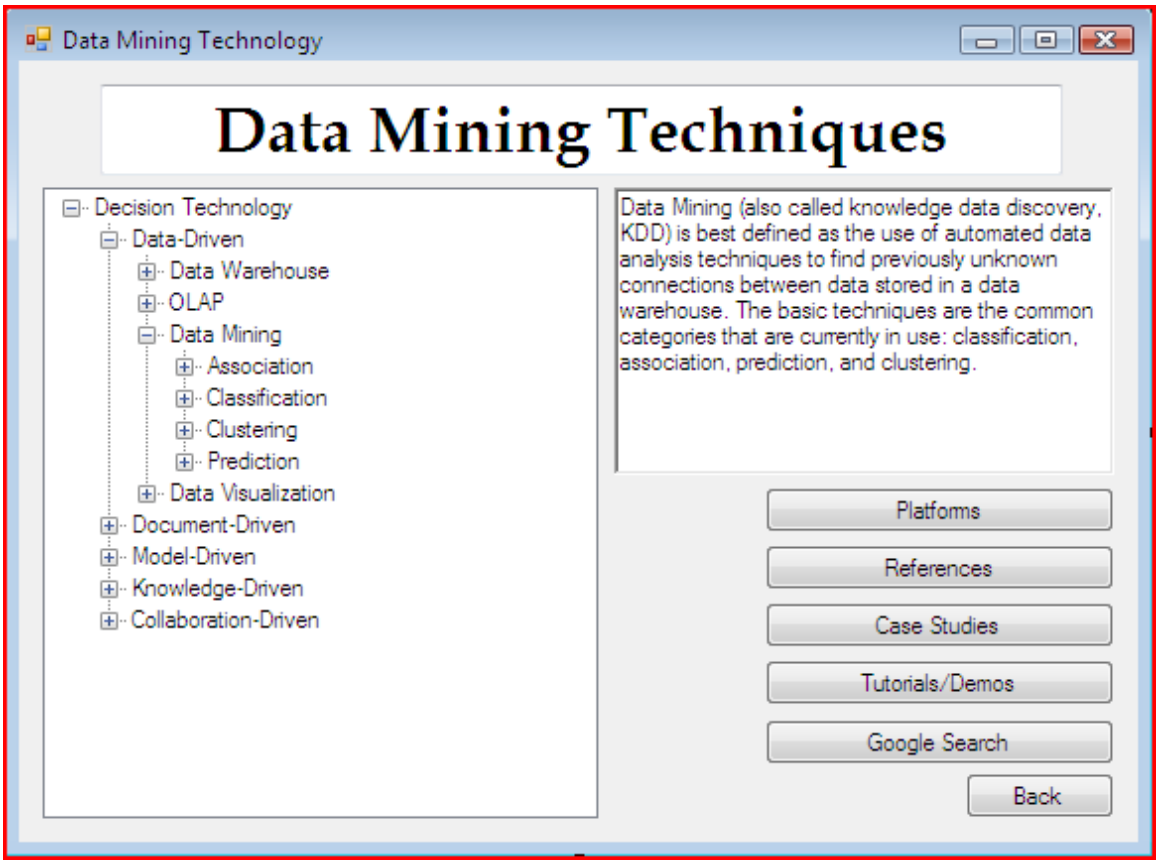

Figure 22. Data Mining Techniques Screen.

Using the mouse-over feature, the user can determine which data mining category is most likely to contain the Decision Tree technique. In this case, the Prediction Technique is the appropriate category, and the associated screen is shown in Figure 23.

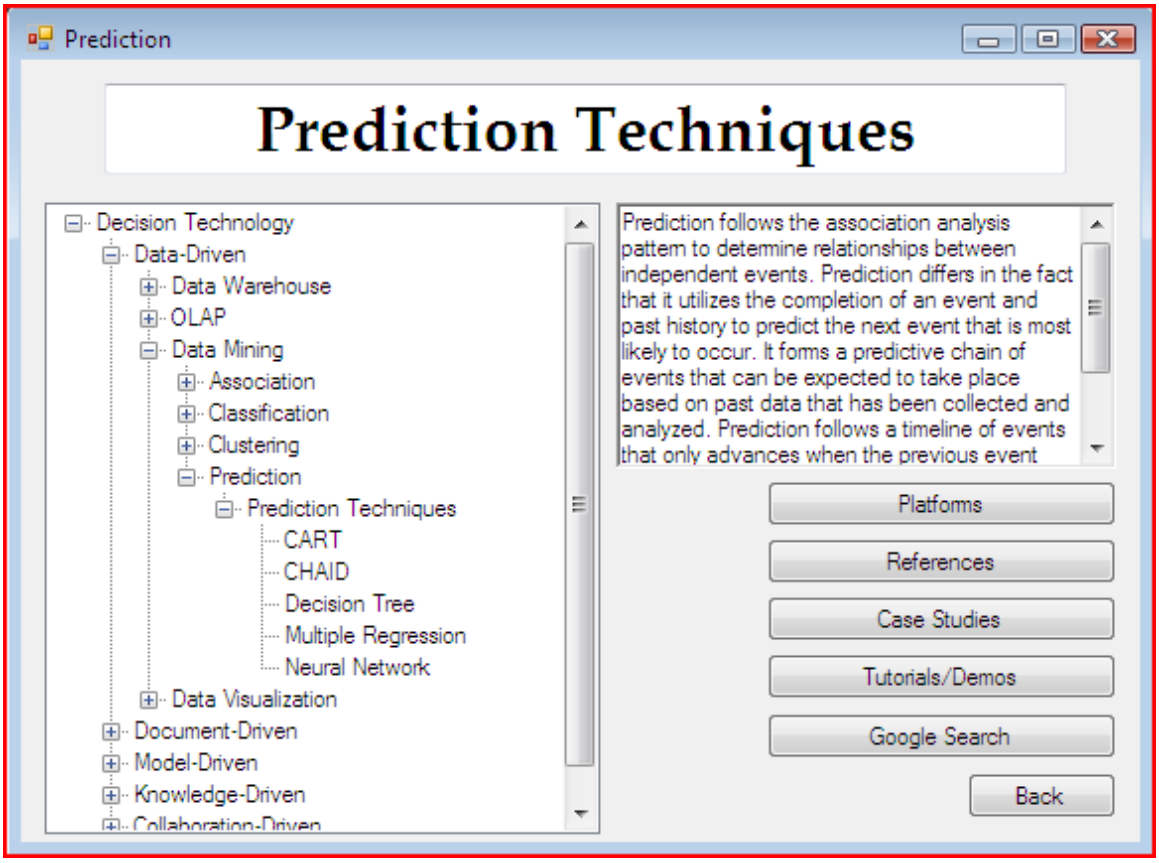

Figure 23. Prediction Techniques Screen.

The user then will navigate to the Decision Tree screen (Figure 24). This screen, as the others, provides the user with a brief definition of Decision Trees, the Windows Explorer navigation tree, and buttons to perform select tasks.

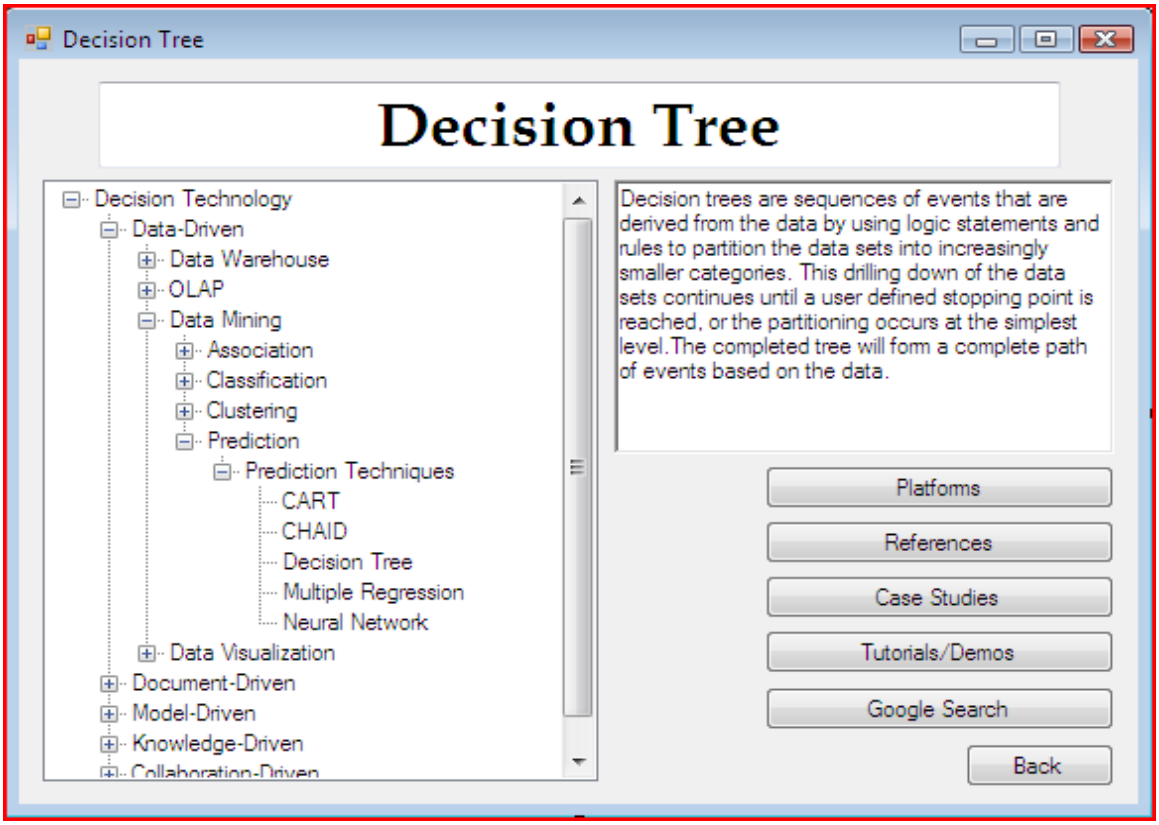

Figure 24. Decision Tree Screen.

From this screen, selecting the *Platforms* button shows the platforms available at NPS to perform Decision Tree analysis, namely SPSS Clementine and Megaputer PolyAnalyst. The SQL query linked to the *Platforms* button would be:

```
SELECT [Decision Technology].DT_Platform_Software_System,
DM_Technique.Technique_Name
FROM DM_Technique INNER JOIN (([Data Mining_DT] INNER JOIN
[Data Mining_DT_Supports] ON [Data Mining_DT].ID_ = [Data
Mining_DT_Supports].[Data Mining_DT_ID_FK]) INNER JOIN
([Decision Technology] INNER JOIN (DataDriven_DT INNER JOIN
DataDriven_DT_DD_Category ON DataDriven_DT.ID_ =
DataDriven_DT_DD_Category.DataDriven_DT_ID_FK) ON [Decision
Technology].ID_ = DataDriven_DT.[Data_DT_Id_Decision
Technology_ID_FK]) ON [Data Mining_DT].ID_ =
DataDriven_DT_DD_Category.[Data Mining_DT_ID_FK]) ON
```

```
DM_Technique.ID_ = [Data
Mining_DT_Supports].DM_Technique_ID_FK
WHERE (((DM_Technique.Technique_Name)="decision tree"));
```

Figure 25 shows the screen as it would appear to the user once the *Platforms* button was selected.

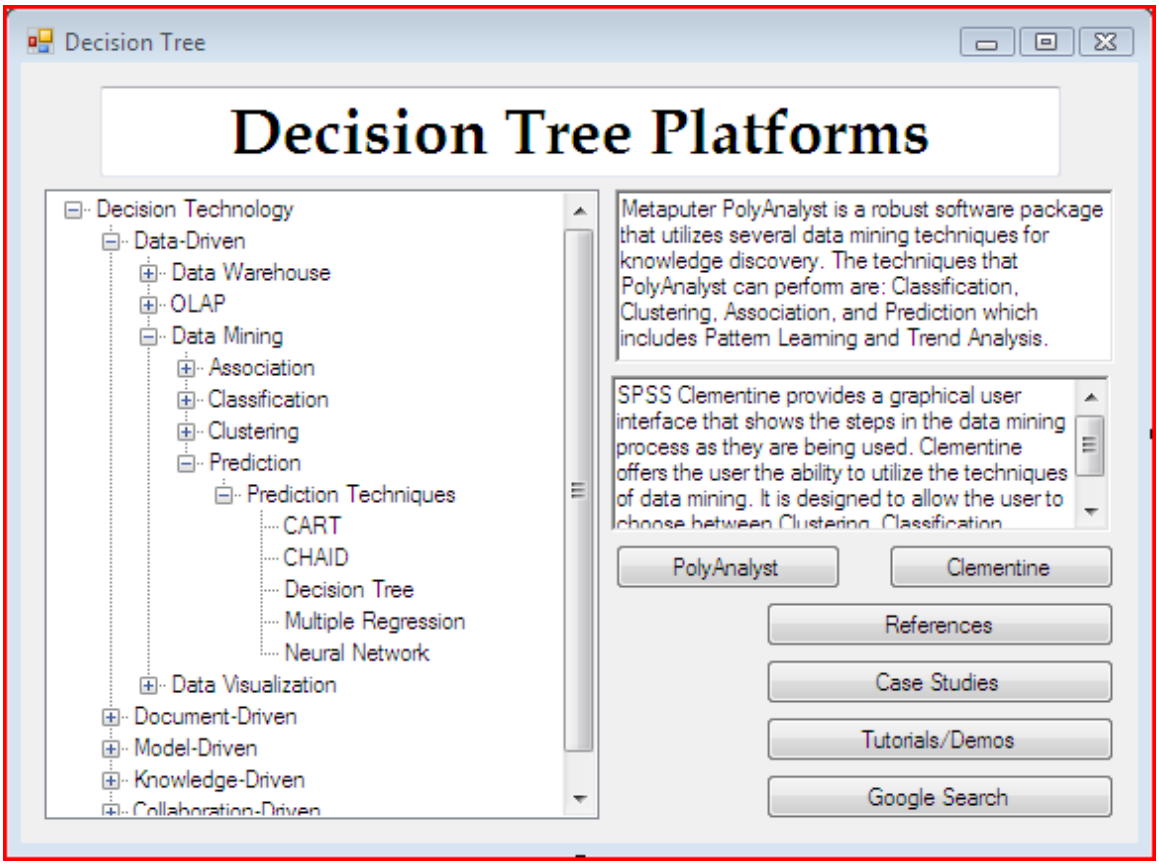

Figure 25. Decision Tree Platforms Screen.

Figure 26 and Figure 27 show the PolyAnalyst and Clementine start-up screens respectfully. From there the user can begin the decision tree and data entry.

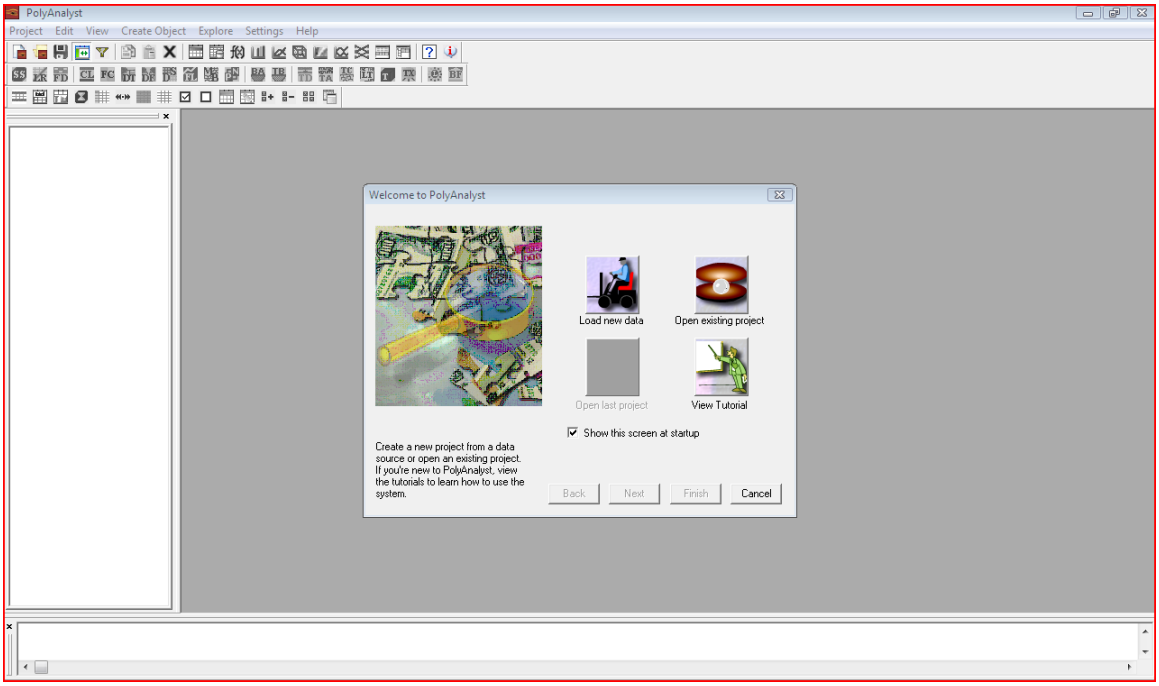

Figure 26. Megaputer PolyAnalyst Start-up Screen.

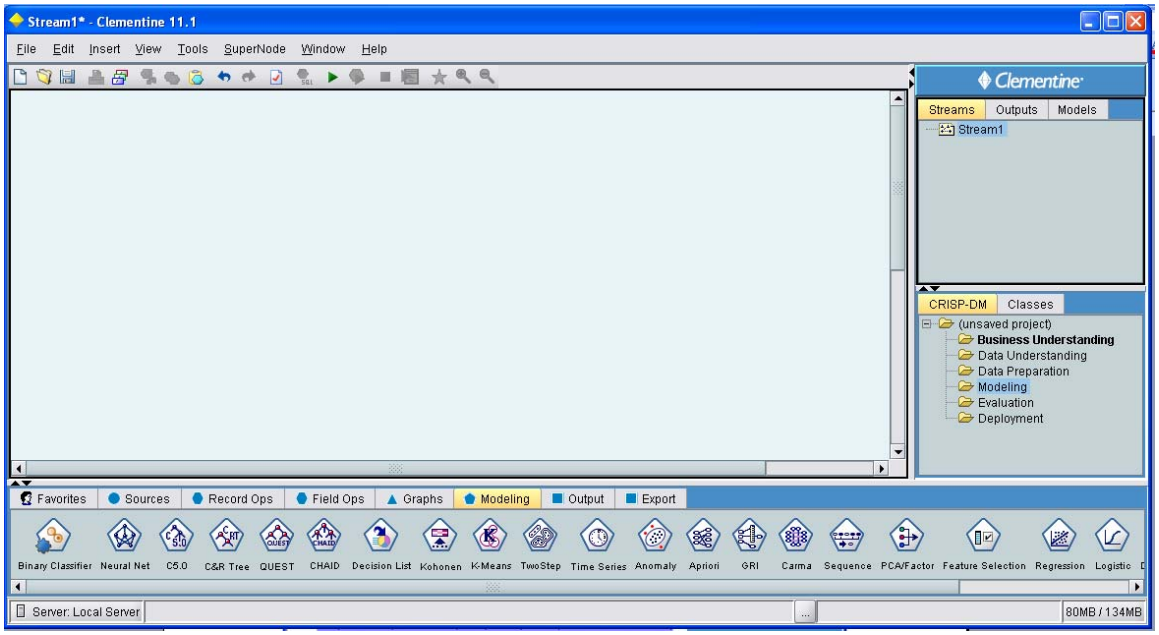

Figure 27. Clementine Start-up Screen.

Also from the screen in Figure 24, selecting the *Case Studies* button shows the Case Studies available at NPS that

have utilized Decision Tree analysis. The SQL query that would be linked to the *Case Studies* button would be:

```
SELECT [Case Study].Description AS [Case Study_Description],
DM_Technique.Description AS DM_Technique_Description
FROM DM_Technique INNER JOIN (([Case Study] INNER JOIN [Case
Study_Model] ON [Case Study].ID_ = [Case Study_Model].[Case
Study_ID_FK]) INNER JOIN [Case Study_Model_Process] ON [Case
Study_Model].ID_ = [Case Study_Model_Process].[Case
Study_Model_ID_FK]) ON DM_Technique.ID_ = [Case
Study_Model_Process].DM_Technique_ID_FK;
WHERE (((DM_Technique.Technique_Name)="decision tree"));
```

The remaining buttons on the screen (*Reference*, *Tutorials/Demo*, and *Google Search*) would be attached to a hyperlink that would open automatically open an Internet Explorer window and navigate to the appropriate web page. Examples of the hyperlinks are:

- "Reference": one might see links such as http://infogoal.com/dmc/dmcdwh.htm
- "Tutorials/Demo": a link would point to a location on the server that held the software platform. For example, the Megaputer PolyAnalyst Tutorial would be hyperlinked to the following file location:

  C:/Program%20Files/Megaputer%20Intelligence/PolyAnalyst%25.0/Help/toc.html
- Google: this would redirect the user to http://www.google.com, or alternatively one might embed the Google search software within the IDTE environment so the search engine could be invoked at any stage of the process.

D. FUTURE ENHANCEMENTS

The IDTE can be enhanced with additional features as requirements are identified. One particularly desirable feature would provide a local, search-based capability. This enhancement would place a search bar on the navigation pages that allows the user to specify search criteria for identifying screens from the Decision Technology Explorer outline. This would provide an alternative to the hierarchical navigation which we've designed. So if a user types "decision trees" into the search engine, the IDTE would navigate directly to the Decision Tree Page shown in Figure 24 without having to navigate through the screens in-between.

E. SUMMARY

In this chapter we developed a conceptual flowchart for a preliminary version of the IDTE user interface, focusing upon the data-driven and data mining decision technologies. The interface was designed from the decision technology taxonomy and related database schema developed in the previous two chapters respectively. The flowchart yields a simple, yet consistent interface which provides users several different perspectives for learning about and using available decision technologies at NPS.

THIS PAGE INTENTIONALLY LEFT BLANK

V. SUMMARY

Decision support technologies have remained individualistic in nature. The ability to access and integrate a wide range of such technologies in an Integrated Decision Technology Environment can potentially increase a user's ability to create more complex decision support projects. The IDTE will allow the user to learn, access and use available decision technologies quickly and easily.

Data-driven DSS provide decision-makers with the capability to store and sort through potentially vast amounts of data by leveraging data warehousing and data-mining. These data-oriented decision technologies can assist decision-makers in making better and more informed decisions in shorter durations of time.

This research has focused upon data-driven decision technologies and how they can be integrated into an overall IDTE. In the process of identifying data-driven technology requirements, we first specified a simple taxonomy, based upon their properties and roles. Four categories were identified: association, classification, clustering, and prediction. Next we developed a database schema for storing the relevant data about these technologies including platform data as well as case studies. Finally we designed a simple, yet effective, interface for navigating through the data-driven decision technology universe both at NPS and beyond. SQL commands for populating the various screens of the IDTE interface were provided to show proof of concept.

The thesis clearly showed how an IDTE could be useful to students and researchers at NPS in their ability to

define more complex working models of DSS. The conceptual model for IDTE presents an exciting opportunity for decision technology development, while advancing the opportunity for integration of different decision technologies into a cohesive environment. As additional decision technologies are incorporated into the IDTE, it offers the ability to address ever more complex decision-making challenges.

LIST OF REFERENCES

Berkhin, Pavel. Survey Of Clustering Data Mining Techniques (San Jose: Accrue Software, 2002).

Harris, Shon. All In One CISSP Exam Guide, 3rd ed. (New York: McGraw-Hill, 2005).

Marakas, George M. DSS In The 21st Century, 2d ed. (Upper Saddle River: Prentice Hall, 2003).

Megaputer PolyAnalyst Brochure. 2007. Online. Internet. 12 Feb 2008. Available from http://www.megaputer.com/polyanalyst.php.

Power, Daniel J. DSS: Concepts And Resources For Managers (Westport: Quorum Books, 2002).

Ricardo, Catherine M. Databases Illuminated (Sudbury: Jones and Bartlett, 2004).

SPSS Clementine 11.1 Specification Brochure. 2005. Online. Internet. 12 Feb. 2008. Available from http://www.spss.com/pdfs/CLM11SPClr.pdf.

Tan, Pang-Ning, Michael Steinbach, and Vipin Kumar. Introduction To Data Mining (Massachusetts: Addison-Wesley, 2006).

Turban, Efraim, and Jay E. Aronson. DSS and Intelligent Systems, 5th ed. (Upper Saddle River: Prentice Hall, 1998).

THIS PAGE INTENTIONALLY LEFT BLANK

INITIAL DISTRIBUTION LIST

1. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California

2. Professor Dan Dolk
   Department of Information Sciences
   Naval Postgraduate School
   Monterey, California

3. Albert “Buddy” Barreto
   Department of Information Sciences
   Naval Postgraduate School
   Monterey, California

4. Brian Hargrave
   Monterey, California